DOCUMENT RESUME

ED 399 808 FL 024 135

AUTHOR Green, Georgia M.
TITLE The Right Tool for the Job: Techniques for Analysis of Natural Language Use.
PUB DATE 95
NOTE 18p.; In: Pragmatics and Language Learning. Monograph Series Volume 6; see FL 024 134.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Collection; *Discourse Analysis; Language Patterns; *Language Research; Languages; Linguistic Theory; *Natural Language Processing; *Pragmatics; *Research Methodology

ABSTRACT

A variety of techniques for collecting and analyzing information about the natural use of natural languages is surveyed, emphasizing the importance of recognizing the properties of a research task that make a given technique more or less suitable to it rather than comparing techniques globally and ranking them absolutely. An initial goal is to characterize the sorts of tasks that are involved in research on language use. Language use is viewed as a complex interplay of: (1) properties of the linguistic system; (2) properties of the language users; and (3) properties of the society within which a speaker elects to use a particular language on a particular occasion. These factors then determine the kinds of questions to be asked about language use. Advantages and disadvantages of different sources of language for analysis are considered, including large corpora, audio recordings, and interactive techniques for eliciting language use. It is concluded that in most cases, the best way to test a hypothesis about natural language use is to test it in every relevant way possible. Contains 32 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 399 808

THE RIGHT TOOL FOR THE JOB: TECHNIQUES FOR ANALYSIS OF NATURAL LANGUAGE USE*

Georgia M. Green

This article[1] surveys a variety of techniques for collecting and analyzing information about the natural use of natural languages. It emphasizes the importance of recognizing the properties of a task that make a given technique more or less suitable for it, rather than comparing techniques more globally and attempting to rank them in any absolute fashion. Thus, an initial goal is to characterize the sorts of tasks that are involved in research on language use. The conclusion is that no single tool is good for everything (the best screwdriver makes a lousy hammer, and a worse saw). It follows from this that disputes about which tools are "best" boil down to questions about the values and assumptions that have implications for which tasks are important.

INTRODUCTION

When Searle (1969) popularized the notion of speech acts twenty-five years ago, it triggered an unprecedented explosion of interest in the study of language use. Today, groups large and small study in more or less systematic ways principles of language use governing large and small classes of linguistic acts (ranging in scope from whole texts and interactions to monosyllabic interjections), at finer and grosser degrees of granularity. The techniques that they have developed for answering the myriad of questions that researchers are stirred to ask are so diverse that many researchers are unfamiliar with more than a few. There is a concern that the growing focus on techniques threatens to fractionate the field of pragmatics research into warring camps of paranoid cults, each believing they have found the One True Way to investigate questions of language use.

It turns out (hardly surprisingly) that belief in One True Way entails circumscribing the set of questions that define the field to just those that can be answered by that One True Way, and naturally enough, different ideas about the One True Way determine sets of questions that are not congruent with each other.

The purpose of this article is therefore frankly ecumenical: all of the techniques discussed are valuable, though they are not all valuable for the same task. Thus, the first task is to describe the character of pragmatics research in terms of a broad view of the domain of language use, and to outline the sorts of questions one might ask, and the sorts of information required to answer them. In this context, it is easier to see what sort of technique is most useful for getting a handle on the specific sort of information that is desired.

FL024135

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Lawrence F Burton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

THE TERRITORY

Since what tasks are required in the study of pragmatics is a function of the knowledge sought, I begin by characterizing what kinds of knowledge we seek as investigators of pragmatics.

Pragmatics is about language USE, so the goal here must be to learn when and why particular linguistic forms are USED. We don't know, a priori, if answers to these questions are to be found by understanding such questions in terms of the structure of abstractions treated as formal structures, or in terms of assumptions and motives or language users, it will be important to frame the goals so as not to preclude any possible answers. (I do have a bias here, but I will try to keep its interference at a minimum.) Suppose we phrase the questions this way:

- What are the significant classes of uses of language? (i.e., what are the identifying properties of the various phenomena of interest?)
- What governs the distribution of members (or tokens) of the classes (or types)? That is, under what circumstances do we expect to find an instance of one type rather than another?

THE DOMAIN OF LANGUAGE USE

There are so many aspects of language use that any single classification imposes a distortion on the analysis of research issues. Consequently, to organize the discussion of techniques in terms of tasks, I will try to elucidate the relations among the subfields of pragmatics research in terms of a cross-cutting multiple-perspective approach.

Language use in context is, evidently by definition, a complex function of three things: 1) properties of the linguistic system, abstracted away from particular occasions of use, 2) properties of language users, independent of what language, if any, they choose to use for communicating and for affecting and effecting events, and 3) properties of the societies within which a particular speaker elects to use a particular language on a particular occasion of use.

Linguistic Systems

Languages provide means of structuring information in two quite distinct ways. First of all, individual words and their relations to other words provide conceptual categories according to which objects and events perceived to be real might be classified and treated as the same or different. This alone is enough to keep an army of pragmaticists busy, exploring the programme outlined by Nunberg (1978) in *The Pragmatics of reference* to explain how people unconsciously gauge what set of referents might be (expected to be) intended for a particular word on an occasion of use. The problem is illustrated in (1), but that is only the beginning.

(1a) Racceons are herbivorous.
(1b) Raccoons knocked over a garbage can and had a pizza party on my porch last night.
(1c) Raccoon has more cholesterol than squirrel.
(1d) Raccoon is warmer than chinchilla.

Of additional interest to pragmaticists is the fact that some words have invariable conventional implicatures (presuppositions) associated with their use. Thus, use of active verbs, as in (2), conventionally implicates a presupposition that the content of their complement is true.

(2) Kim realizes that the Board is corrupt.

The use of a variety of other words and morphemes (especially, say, pronouns and honorific affixes) is associated with presuppositions about the social relations among speech act participants and referents of expressions in utterances.[2] Still others, typically adverbs (*consequently*, *however*, *moreover*, *thus*, *now*, *so*) and conjunctions (*but*, *since*) and particles (*well*, *why*, *like*, *OK*, *y'know*) give information about how (the speaker believes) one part of the discourse relates to another, or how an utterance relates to the addressee's attitude or belief system.

Languages also have constraints on how words go together to make phrases, and particular phrasal constructions may be of interest to pragmaticists as well, because their use may (sometimes or always) imply information above and beyond what is predictable from information in the constituent parts and their semantic relation to each other. For example, in many languages, using a passive construction implies a belief that the event described significantly affected the referent of the passive subject (often adversely). Similarly, if one uses a transitive construction that is truth-conditionally equivalent to a construction that treats its direct object as having some other grammatical relation (as in Raising and Dative Alternation sentences), it often implies a belief that interactive potential exists between the referents of the subject and direct object (Green, 1974; Postal, 1974), as in (3) as compared to (4).

(3a) Dan won Jane a gold medal.
(3b) Tracy expected Al to sleep late.

(4a) Dan won a gold medal for Jane.
(4b) Tracy expected that Al would sleep late.

There are shelves of literature just from the last twenty-five years on the specific pragmatic implications of the use of such constructions, going beyond Passive, Raising, and the Dative Alternation, and past Extraposition, Topicalization, Reflexivization, and Inversion to include practically every construction that has been given a name, and many that haven't; the Construction Grammar being developed at Berkeley particularly fosters attention to such matters. It is probably safe to speculate that use conditions exist in every language for any construction whose content could be conveyed in a less complex construction, following principles of contrast elucidated in work by Horn (1984, 1989).

BEST COPY AVAILABLE

Language Users

The properties of language users---speakers and addressees and overhearers---that figure into the description of how a bit of language is appropriately used are their beliefs and intentions. It is not their actual social status or their relative authority (in any sense of the word), but their beliefs about such things and about each other's beliefs about such things that make a difference. Similarly, simple intentions aren't really relevant either. What counts is not an intention that an addressee do some particular thing, but an intention to get her to recognize that the speaker intends that the uttering of some particular expression will get her to recognize that the speaker wants her to do that thing. Thus, the relevant beliefs are 1) beliefs about objects and events in the real world (or any hypothetical world defined in the course of the discourse), 2) beliefs about what interpersonal behaviors are valued by the culture, and what beliefs are routinely ascribed to all normal adult members of the culture, and 3) beliefs about what things have been referred to in the ongoing discourse (however defined) and what has been said about those things and how the presentation of that information has been structured. Some or all of these are involved in the use of pronouns and other deictic and indexical expressions, in the choice of register, and in the negotiation of turn-taking. Finally, the set of relevant beliefs also includes, of course, beliefs about one's interlocutor's beliefs about all of these. The relevant intentions are intentions to change the world, ordinarily by affecting belief states or intentions of an addressee (which already changes the world), typically with an eye to getting her to eventually do something, which will change the world in additional ways. This characterization is intended to cover every conceivable speech act at every imaginable level, from acts of reference and predication to statements and questions and directives to promises and requests and apologies accomplished by utterance of them, to insults and compliments conversationally implicated by such utterances, to third- or fourth-order demonstrations of empathy or politeness that such conversational implicatures might effectuate, and so on. Conversational implicatures arise from the assumption that it is reasonable (under the particular circumstances of the speech event in question) to expect the addressee to infer that the speaker intended the addressee to recognize the speaker's intention in uttering whatever she or he uttered from the fact that the speaker uttered it. Because conversational implicature is based on inferring intentions for actions generally, not just linguistic actions, it is a function of human behavior generally, rather than being something specifically linguistic.[3] As researchers, we face the challenge of determining which of the propositions that go into an implicature are universal, which are culture-specific, and which are linked to specific bits of language.

Societal Aspects

The third dimension of language use that imposes a classification on instances is a projection of the fact that languages do not exist in the abstract, but are associated with particular societies. More specifically, it follows from the fact that languages are systems of CONVENTIONS, the acceptance of which presupposes--one might even say, defines--a society. Conventional aspects of language thus include 1) the conventions of form and of the form-meaning correspondence that constitute the grammar of the language, 2) the conventional beliefs and values that constitute the shared culture of the society, and 3) the conventions of language use (Morgan, 1978), which are language-specific, and yet not strictly part of the grammar of the language. These latter include the enumeration of various

speech-act formulae (e.g., greetings, curses), idioms, and conventionalized (short-circuited) implicature, as well as other rhetorical conventions of the culture, addressing, for example, discourse structure, and modes of indicating it. (Of course, there is no claim here that the conventional beliefs and values, and the conventions of form have to converge in defining a single society; only that at the time of a particular instance of language use, the speaker behaves as if she recognizes that she is operating under the relevant societal conventions in all cases.)

Within the structure provided by these conventions, we all know that individuals make creative use of (conventionally) fixed resources. Speakers (depending on their personality, their wit, and their wont) exercise their creativity in constructing referential terms (through, for example, compounding and metaphor), in choosing to communicate less directly via conversational implicature, and in constructing texts and discourses such as jokes, arguments, explanations, and narratives. For example, a speaker may choose to use an agentless passive to implicate that the identity of the agent of an act referred to is unknown, or irrelevant, or unrevealable. Or she may not mean it to implicate anything at all. Interpreting and disambiguating utterances and negotiating turn-taking all involve modelling one's interlocutor's model of the discourse, and coordinating contributions according to a model believed to be shared.

DIFFERENT KINDS OF QUESTIONS

Just as these three dimensions cross-classify familiar instances of language use, so do they determine different kinds of questions about the deployment of particular forms or types of language use.

Questions About Language Users

Perhaps the most basic research questions in pragmatics have to do with language users and their states of mind on the occasion of use. Why does saying X produce a different effect than saying Y? For example:

Well, vs. *Why*,
Can you VP vs. *Are you able to VP*
tu vs. *vous*

Why does X say Y? What is the purpose of saying Y? What characterizes the occasions of use on which Y is used? Hypotheses about the answers to given questions might involve

properties of the speaker,
properties of the addressee,
properties of the relation between speaker and addressee, as well as
beliefs and/or
goals of the speaker.

Questions About Culture

Higher-order questions of language use may be linked to culture and societal convention. For example: Is Y (some type of language-use) defined the same way in Culture A and Culture B? Are the subtypes of Y the same in both cultures? If so, are the principles governing their distribution the same? Do they have the same relative frequency? If not, to what should the difference be attributed? Are they used for the same purposes? Are they used in the same contexts? (This is probably the same question, in different clothing.) What makes using Y polite/rude? Is it always polite/rude? Why is it hard for speakers of Language A to learn to use X appropriately when speaking Language B?

Structural Questions

Finally, questions may be framed, wholly or largely, in terms of structures or forms: What properties of the structure of a discourse determine or affect the distribution of form Y or type Y? What structural properties of a form determine or affect its distribution in a discourse? What are the formal or structural subtypes of a particular type of use--i.e., what kinds of linguistic forms can instantiate this use?

SORTS OF TECHNIQUES FOR GETTING ANSWERS

As researchers we obviously don't go directly from identifying a research question to "gathering data" that we expect to bear on its resolution. There is always the intermediate step of latching on to a hypothesis which strikes us as a potential answer, or a set of alternative hypotheses. Once we have a hypothesis, we make a beeline for data that will test it, and either corroborate it, or disconfirm it. (We do this not because we are following some prescribed scientific method, but just because we have operational minds and act rationally.) It matters little where the hypotheses themselves come from--we may regard them as gifts from the muses. What data we want depends on what hypotheses we want to test, and we can't rationally collect data unless we have a hypothesis. If we try, we quickly realize that we don't know how far afield to range, or when we have enough, because we don't know what we're looking for. There is no such thing as theory-independent research. Research is always defined within the framework of some sort of theory (i.e., hypothesis) however generic or underspecified it may be.

Obviously, the kind of information being sought must affect the choice of technique to find it. If you want to know whether words supposedly belonging to the same register distribute the same way across genres, no amount of interviewing speakers, no matter how cleverly, will provide the kind of definitive quantitative data that electronic searches of large, sorted corpora can provide. If you want to know what sorts of beliefs are reflected in the use of some particle, or some special morphology, no amount of studying transcriptions of natural speech, no matter how finely described, will yield an answer, if that is all that is examined. If you want to investigate what beliefs are reflected in the use of some form, you have to be in a position to make testable inferences about beliefs. Because inferences drawn from recorded positive data need to be confirmed by judgements on corresponding negative data, producing a contrast set (a minimal pair) to test whatever hypothesis beckons entails

dealing with constructed data, and with intuitions about hypothetical uses in hypothetical contexts.

"Natural Speech" Only?

Since I've brought up the I-word (*introspection*), perhaps prematurely, something must be said about the value-laden distinction between research that starts with a body of natural speech and research that is organized around discourse segments constructed by an investigator. First off, it is doubtful that this is a very useful distinction to make; it generates all the heat and passion and excessive rhetoric of abortion policy debates (including ideologically-driven disputes over nomenclature--there is an agenda behind the decision to call imagined discourse "artificial" or "constructed"), with only a shadow of the significance.

There are three reasons to consider it a less than useful distinction. First, the same sorts of reasons that motivate a distinction between competence and performance militate against the glorification of The Actually Said (I warned you about it generating excessive rhetoric, didn't I?): it makes it impossible to distinguish between slips of the tongue and intended utterances.

Second, it is not clear that the distinction can be drawn in a useful way. Surreptitious recording of spontaneous speech yields natural data, but, for good or ill, it is considered unethical, and contrary to the guidelines for research on human subjects. Overtly recording natural conversation subjects the data to the charge of experimenter influence: how can we know that people didn't put on airs or become inhibited and talk unnaturally precisely because they knew they were being recorded? There are various large, electronically accessible collections of connected text which are considered "natural language" (i.e., found objects, as it were), but the recorded language is largely written language, that is, planned discourse, and open to the charge of being artificial, and/or artistically manipulated rather than truly natural and spontaneous. Actually, of course, all speech is subject to this charge--we all try to be witty on occasion--so it's not clear that written text (newswriting, exposition, narrative) is necessarily inferior to spoken language. But if written narrative and fictive dialog are valid sources of natural language use, why should it make a difference whether the writer got paid for it, or is a linguist (and therefore, didn't get paid)? There must be bad writers and good writers in both groups. The fact that the language got written down, or the purpose for which it got written is surely too crude a criterion to distinguish the useful from the not so useful.

Finally, all language use is use in context. With natural speech, we know that it was used in a context, but we can never know that we have the relevant information about that context. The factors we bring to bear as analysts in interpreting and/or classifying that particular bit of speech are what we IMAGINE about the speaker's beliefs and intentions regarding the effect of the utterance upon the addressee, whether or not we recognize it either implicitly or explicitly. Thus, there would seem to be no particular significance to knowing whether someone actually said a particular bit of speech, or only might have said it.

BEST COPY AVAILABLE

That said, what follows is an overview of techniques organized along the 'spontaneous-vs.-constructed' dimension, ranging from uses of large, electronically accessible databases of connected discourse, to detailed representations of the physical aspects of speech, to an assortment of clever and useful variations on the Linguist's Creed ("Can you say this?") that have been developed over the past 25 or 30 years.

LANGUAGE USE RESOURCES

Large Corpora

A variety of large, machine-readable corpora of connected English discourse (some of it parsed) are readily accessible through the Oxford Text Archive and the International Computer Archive of Modern English (ICAME) in Norway, the best-known of these corpora being the million-word Brown Corpus of printed English. The Brown corpus of American English (compiled in the 1960s), and the Lancaster-Oslo/Bergen Corpus (LOB, compiled in the 1970s to match the Brown corpus as closely as possible), consist of 500 "randomly"[4] selected 2000-word samples of discourse from fifteen genres (including newswriting, academic prose, science fiction, romantic fiction, skill and hobby instruction, and humor). The London-Lund Corpus of Spoken English, compiled in the 1970s and 1980s, contains half a million words of prosodically transcribed monologue and dialogue, with individual words annotated (or TAGGED) for part-of-speech, and segments identified by speaker demographics (e.g., "female undergraduate, age 20"). Some of the dialogue is face-to-face, and some is recorded surreptitiously. The Scandinavians have been cranking out studies of discourse particles and syntactic constructions from such data for years.[5] A recent book devoted entirely to transcription and coding in discourse research (Edwards and Lampert, 1993: *Talking Data*) allocates an whole chapter (Edwards, 1993) to a survey of electronic corpora and related resources, and it has complete access information for many of these.

In addition to the ready-made corpora, you can roll your own. With an optical scanner, or by transcribing your own audiotapes, you can fashion a corpus tailored to your particular needs. Public language (political speeches, public lectures, radio and TV talk shows (especially call-in shows) can be recorded for personal use (including non-commercial research), and often networks are willing to provide transcripts at low cost. (The transcripts may be fairly primitive, from a linguist's point of view, because they may edit out much language use of interest (such as hesitations and repetitions), but they still save time in the preparation of a more useful transcript from tapes. For that matter, books that are compilations of edited interviews, like Studs Terkel's *Working* (1972) and *Division Street* (1967), are useful for investigating the use of particular words and constructions.)

If you want a tagged or parsed corpus, there are public-domain taggers and parsers.

Uses. The key virtue of machine-readable corpora is that they can be machine-processed to scout the variety of uses of any form that can be specified as a string of characters. They provide easy access to large amounts of relatively unselected data, the searching of which by hand would be prohibitively tedious, and indubitably liable to oversight. The utilities for searching for defined strings of characters that come with practically every word-processing

program can be used to locate every instance of a form that matches the defined string of characters, for example: you know or OK or okay or but or therefore. Concordancing programs that index selected strings and save them to a file with as much surrounding text as you desire are not quite a dime a dozen, but there are several available, and some are free or nearly so.

Parsed corpora, and to a lesser extent, tagged corpora, are very useful for quickly gathering a large sample of instances of a construction type (rather than a particular word) whose use is of interest. Suppose, for example, that you wanted to study the use of the passive. If you tried to collect instances by searching for all words ending in *-ed* (*ed), not only would you have to sift through unwanted nouns (*bed*), verbs (*trembled*), and adjectives (uninhabited), you would miss all the passives with irregular participles, like *rung*, *thought*, *struck*, etc. If you tried to locate relative clauses introduced by *that*, you would have to wade through instances of demonstrative articles and pronouns and complement clauses introduced by *that* as well. If you wanted to survey relative clauses with no introducer (like *the horse I rode*), you'd be plumb out of luck. However, if you can search a parsed corpus for passive verb phrases or relative clauses, finding what you're interested in becomes a lot easier.

As mentioned earlier, the most universal appeal of large, machine-readable corpora is the opportunity they afford for scouting the territory, for getting a glimpse of the variety of contexts in which the form or construction of interest appears. In my experience, this variety is dependably many times greater than the investigator imagines before performing the search.[6] In addition, machine-readable corpora are ideal for researching quantitative properties of texts or text-types, since it is a relatively simple matter to get computers to count instances of things, and compute ratios (say, of definite to indefinite articles, or words per sentence).

Finally, the ability to use the blind-search capacity of computerized interfaces to machine-readable text corpora in order to search for certain sorts of correlations is perhaps of broader interest. Such interfaces (which can be as simple as a text-editor for word-processing) are especially useful for this since they eliminate the need to pre-analyze the corpus and code up every property of potential interest in every segment into machine-interpretable form so that a number-crunching program can look for correlations at a specified level of significance. This is potentially a real boon, because of all the decisions that have to be made in coding a text this way, including even what principles to use to segment it. If any one of them turns out to have been a bad decision, the whole corpus may have to be re-coded before searching can resume. If the properties of actual interest are identifiable as annotations in the corpus, concordancing programs enable searches for

< string > within N < units of text: words/characters/turns... > of < string >

You can search just the relevant segments of the corpus, and do your own tabulating, and feed the numerical results to a statistics program to determine their statistical significance.

One might search for correlations among such machine-accessible properties of text segments as use of particular forms (expressions or constructions), speaker demographics (e.g., age, sex), text genre, among others. Researchers might be interested in correlations

BEST COPY AVAILABLE

of such properties with presumed local goals of speakers, that is, speech act types. Unfortunately, corpora don't come with speech acts tagged, so you're on their own here, and will have to come to grips with the conversational nature of speech acts: often utterances represent several acts at the same time, some relatively direct and some by implicature from those. Thus, stating *I got it for you yesterday* might be an explanation, and by virtue of being an explanation, a rationalization or an excuse, or a refusal.

Of course, finding correlations is probably not so valuable as an end in itself as it is as a source of new hypotheses. Correlations that are reproducible over different sets of data are not just robust results. They are also mysteries to be explained. Any speculation about why X and Y correlate is a new hypothesis to be tested. And of course, correlations can themselves constitute an indirect test of a hypothesis. If form X is claimed to serve some function Y, and it is accepted that Y is a typical function of texts of type Z, then the hypothesis will be corroborated by discovering that X appears more frequently in texts of type Z than in texts of types where Y is not a salient or typical function. Of course, if there are more direct ways to test the hypothesis, then it's only a weak corroboration. Svartvik (1990), Garside, Leech, and Sampson (1987), and Johansson and Stenstroem (1991) offer a variety of perspectives on building and exploiting parsed corpora.

Limitations. Glowing endorsements aside, there are limits to the utility of electronic corpora. For one thing, they offer nothing to the study of usages whose form is not characterizable as strings of characters or representations of syntactic structures, and typically research on particular sorts of speech acts falls into this category, since any speech act can be accomplished by an unlimited variety of forms--typically, a few performative forms (like *I apologize*), a larger number of forms which are used to accomplish the act by conventionalized conversational implicature (like *I have to apologize*, *I'm afraid I have to apologize*), and an unlimited number of forms the utterance of which conversationally implicates the goals and attitudes that characterize that particular speech act, for example, in the case of apologies: *I really feel awful about that, I hope you'll forgive me, It was really thoughtless of me to do that, Tell me what I can do to fix things, How can I make it up to you?* When cultural values make creativity of expression a hallmark of sincerity, it becomes virtually impossible to specify the class of utterances which are used to perform any speech act in terms of the words and syntax employed.

Second, while I've made parsed corpora sound like the best thing since microwave ovens, they are only as good as the parser and the grammar behind them. There are a lot of pretty good parsers around, both human and inhuman, but there isn't much in the way of comprehensive grammars for automated parsers to use. The best of the grammars only cover a small fragment of the language they describe. For example, I have yet to find a machine-readable grammar of English that can parse with any semblance of utility focus-inversion sentences like *In the corner lay a tattered paperback*, a construction whose uses have engaged me for over 20 years.

Moreover, if an automated parser doesn't just fail to provide a parse when it encounters a sentence with a construction not specifically described in its grammar, the off-the-shelf parse which its default mechanism may provide may be an off-the-wall parse, for example,

treating a sequence of words it can't parse otherwise as representing a noun-noun compound like *employment resources director hiring*.

Even hand-corrected parsed corpora like the Penn Treebank (Santorini, 1990) are only as good as the linguists who did the correcting, and the tag-set and the grammatical theory that they used to do it.

Finally, it is unclear how to interpret statistical descriptions of language use within a corpus, because it is unclear what the sample of texts would be a representative sample of. If eighty percent of the samples came from, say, highly educated white Americans, what does that make the corpus a sample of? If the class, occupation, sex, age, education, dialect (etc.) of the speakers is unrecorded, how do you guess what domain any results generalize to? Worse, because the ways in which and the purposes for which language can be used are unlimited, it is impossible to say how representative of anything ANY particular sample is.

Audio Recordings

Electronically accessible audio-recordings, and detailed, prosodically annotated transcriptions of audio recordings are useful for researching how the prosodic properties of an utterance of an expression correlate with its distribution. A variety of tool kits (some affordable and publically available[7]) exist for electronically analyzing waveforms in audiorecordings, and more seem to be appearing daily.

Uses. Databases of audio-recordings can be used to investigate whether variation in, say, intonation or timing, correlates with any pragmatically relevant aspects of language use, i.e., with demographic properties of speakers, with structural properties of discourses, or with presumed or plausible local goals of speakers. Thus, a person interested in the uses of a discourse particle might discover that it had two very different pronunciations, one with a full vowel that occurred before a pause, and one with a reduced vowel that occurred with what they used to call close juncture, and so be motivated to look for correlations of pronunciation with demographics, discourse function, or speaker attitude.

Limitations. Of course, prosodic analysis is only useful when particular instances of an utterance can be compared with baseline and range information for its speaker. If you don't know what a speaker's normal pitch range and speech rate are, and how both vary with utterance length, you don't know whether a particular piece of an utterance has extra high or low pitch, or whether a brief period of silence should count as a pause.

In addition, if pronunciations are to be correlated with local goals (whether interpersonal or discourse structural), you have to have a way of investigating goals (which presumably reside in the minds of speakers). Actual speakers of recorded discourse usually aren't available to tell you about the goals they had in saying each utterance the way they said it, and rarely can articulate those goals anyway, even if they CAN remember what they were and aren't inclined to misrepresent them in order to protect their self-image. This doesn't mean that goals cannot be researched empirically, only that more subtle techniques are needed to get at them.

Interactive Techniques--Oh, See Can You Say...

Researchers have developed a variety of techniques for testing hypotheses about correlations between usage and intensional attitudes. One way to learn how people's linguistic behavior is affected by having particular sorts of beliefs and goals is to look for a situation where it is reasonable to assume people have those beliefs and goals, and watch to see what they do. For example, you can hang around train stations and eavesdrop on people buying tickets and seeking gate information (Horrigan, 1977; Allen, 1979), or at McDonald's and record them ordering hamburgers (Merritt, 1976). You can record advising sessions or conflict resolution appointments and the like. Or, you can put volunteers in a situation where you can be reasonably certain of their relevant beliefs and goals, and see what they do in that situation. You can show people a videotape and get them to describe it to you, as Chafe's group did in Berkeley in the 1970s with the Pear stories (Chafe, 1980). Or you can set up situations where people have to talk to each other, say, friends describing frightening or embarrassing experiences, or the layouts of their first apartments. In the 1970s, one research project persuaded volunteers to be recorded coaching other volunteers in assembling a toy pump (Grosz, 1977).

Naturally, the more variables that can be controlled, the better the likelihood of getting a meaningful analysis of variation observed in the data. Thus, in the pump assembly experiment, the task was to assist an unseen person in assembling a toy pump, and volunteers were either experts (who had assembled pumps and knew names for the parts and the tools), or novices. Participants knew whether they were talking to other experts or to novices. In the Pear Story experiments, not only was the film plotless, in that the sequence of events filmed was intended not to imply any particular connecting relations among the events, volunteers were allowed, in retelling the events to the experimenters, to make any assumptions they wanted about the level of detail that would be relevant, and about the familiarity of the audience with the events and objects depicted.[8]

At the same time, the more natural the set-up situation is, the more likely that inferences about speech produced in that situation will provide information about the normal use of natural language. Insofar as Pear Story subjects didn't have an internally motivated purpose for retelling what they had seen, the utility of their narratives is diminished by the fact that describing a sequence of events of uncertain import to someone who may or may not be familiar with them is a relatively unnatural act. In addition, to the extent that the respondents inferred or invented a more particular motivating purpose, that constitutes an additional, uncontrolled, source of variation in the data.

Another way of testing hypotheses about the connection between the use of linguistic forms and speaker attitudes uses judgement tasks, in more sophisticated versions of tried-and-true armchair methods. In a series of questions administered as a survey interview, the researcher describes a hypothetical situation where the relevant beliefs and attitudes are explicitly attributed to a speaker, and asks the volunteers what they would expect the speaker to say, or whether some particular response would be more appropriate, or more likely than some others to succeed in accomplishing the goals attributed to the speaker in that situation.

A variation on these methods is to describe a hypothetical situation, and ask volunteers what THEY would say in that situation, or what other people would say, or what other people would say the speaker SHOULD say. Or the researcher might persuade pairs of volunteers to act out the situation, assigning particular roles to each player.

Another well-researched variation, more convenient for mass administration via a written questionnaire, is to describe that situation, and set up a dialogue which respondents are supposed to complete as one of the participants. Such Discourse Completion Tasks (DCTs) may be open-ended (fill-in-the-blank) (Blum-Kulka, House, and Kasper, 1989), or multiple choice (Rose, 1992), and respondents may be asked for rankings of possible answers (Hill et al., 1986; Ide et al., 1992). The advantage of written surveys is that they can be administered efficiently, and a large amount of data can be amassed quickly. The disadvantage is that because it is more tiring to write than to speak, responses are likely to be shorter, and may be less carefully considered. Oral interviews, with responses tape-recorded, take longer to administer, and require an investment of time and effort to transcribe, but the results are more likely to contain richer responses--ones which consider the question in greater depth, and from multiple perspectives. They're probably the best way to test reactions to pragmatic minimal pairs, where the context is held constant, and the difference is just the presence or absence of the form in question, or where the form is held constant, and the context is varied minimally in some relevant way. At the same time, transcripts of oral interviews will contain more false starts and vacillation, which may make coding the responses more difficult.

If respondents are interviewed in groups of close friends, one person's contribution may trigger discussion which exposes new dimensions to the issues under investigation. The other side of the coin is that if there is an exceptionally strong personality in the bunch, her responses may inhibit the others from expressing their true opinions.

Asking people about their own behavior in hypothetical situations would seem to enable getting information from the most direct source. But this information can be skewed by people whose quirks of personality prompt them to behave in atypical and idiosyncratic ways. It can also be skewed if the respondent describes not what she believes she actually would do or say, but what she believes she ought to do or say, what she believes society expects of her.

On the other hand, sometimes information about what the respondent thinks she ought to do, or what (she thinks) other people would do is more informative because it gets more directly at the relevant normal beliefs of the culture--what people believe other people believe everyone believes, and this may be more helpful in characterizing the shared beliefs or conventions that govern usage than even accurate reports of idiosyncratically-governed hypothetical behavior.

Researchers who want to know, for pedagogical purposes, how some sort of language use differs across cultures, as well as researchers interested in discovering what universal principles govern language behavior, have an interest in comparing results of the same survey conducted in different cultures. This is not as easy as it sounds. To have genuinely comparable results, the subject pools must be comparable, and it is naive to think that

BEST COPY AVAILABLE

matching them for age, sex, and education will be enough. A 22-year-old American college student in the U.S. may be a relatively autonomous person living independently of her parents, while her 22-year-old counterpart in another culture may be dependent on her parents for shelter and sustenance. This means that questions about resolving conflicts with a non-family-member with whom the respondent shares living quarters may be entirely beyond the experience of some respondents. Consequently, answers to questions that are superficially the same may not provide comparable information.

Complicating matters further, the situations described in questionnaires and role-playing protocols may be perceived very differently in the different cultures. For example, a request for some particular assistance from a stranger or a social superior that seems unremarkable in one culture might be unthinkable in another.

Although it is probably safe to assume that if something can go wrong, it will, one way to minimize getting non-comparable responses is to include in the research team a member of the other culture who shares not only your understanding of the hypothesis being tested, but also the details of how each projected answer to each questionnaire item tests that hypothesis.[9]

Regardless of where the initial data come from about what people say, or say they would say, or say other people would say, in a given situation, it is often useful to interview respondents about what has been recorded. This can take the form of a play-back protocol where volunteer and researcher review the recorded behavior or questionnaire response, and the volunteer answers questions about why she said this, or what went through her mind when the other person said that. Granted that such responses might be self-serving and so need to be taken with a grain of salt, they can nonetheless provide not only unsolicited corroboration for a hypothesis, but also insight into previously unimagined factors affecting the choice to use some form, which will motivate revising the hypothesis, or the research design, in order to better test the hypothesis.

CONCLUSION

Considering the outline of pragmatics research described here, not much more has to be said about determining which techniques are suited to answering which questions. The hard part of research, as always, is figuring out which questions need to be answered, and being able to roll with the punches, and adjust the question, and the means for testing it as preliminary results reveal more about the domain of inquiry.

As for the practical matter of figuring out the best way to test some particular hypothesis, the best way to test it is to test it in every relevant way possible: questionnaires, interviews, large-scale electronic searches, analysis of natural behavior. Yes, it's an investment, but it is bound to be a worthwhile one; if the results of the tests are not all consistent, you still learn something: namely, that the tests are not all testing the same thing. Figuring out why not, and how to remedy the situation is bound to teach us more about the hypothesis, more about the domain under scrutiny, and more about investigative techniques. How can anyone argue with that?

THE AUTHOR

Georgia M. Green is Professor of Linguistics at the University of Illinois at Urbana-Champaign. Her research focusses on natural language understanding and syntactic theory.

NOTES

*This work was supported in part by the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

[1]These remarks were presented, in slightly different form, to a plenary session of the Eighth Annual International Conference on Pragmatics and Language Learning, held March 31-April 2, 1994 at the University of Illinois in Urbana-Champaign. A number of papers at that conference addressed more specifically some of the particular issues raised here.

[2]For example, the use of referent honorifics is often described in terms of a presupposition that the speaker respects the referent. This is an oversimplification, of course. The speaker doesn't actually have to respect the reference (as represented in such a first-order condition). A more accurate representation of the condition is similar to the conditions Nunberg (1978) described for referential terms, in that it refers to normal beliefs about use of the form: the speaker must believe that (the addressee believes that) it is normally believed that use of the term implies that the speaker respects the referent. This makes the proposition that the speaker respects the referent a conversational implicature of the use of the honorific (cf. Green, 1992). Conversational implicature is addressed in more detail in section on Language Users.

[3]See Green 1993.

[4]The samples were randomly selected from texts that must have been selected on some arbitrary basis; human decisions have to have been involved in determining the categories and the number of samples from each category, and in selecting the texts from which samples might be randomly selected.

[5]Meijs 1987, Tottie and Baecklund 1986, Aarts and Meijs 1990, Greenbaum, Leech and Svartvik 1980 are representative, and Altenberg 1991 provides a comprehensive biography up to 1990.

[6]Fillmore (1991) describes the experience in detail.

[7]For example, the CECIL system for computerized extraction of components of intonation in language produced by the Summer Institute of Linguistics.

[8]Subjects were asked, after being informed that the purpose of the interview was to study how people talk about things they've experienced, to "tell what happened in the movie" to an interviewer who claimed not to have seen it (Chafe, 1980: xiv-xv). It is hard to know how credible the claim of unfamiliarity would be in this context.

[9]In 1993, Lancaster University in England hosted a seminar on introspection in applied linguistics research which covered a lot of these issues, and more, but unfortunately the proceedings are not being published.

REFERENCES

Aarts, J. & W. Meijs. (1990) Theory and practice in corpus linguistics. Amsterdam: Rodopi.

Allen, J.F. (1979) A plan-based approach to speech act recognition. Technical Report 121, Department of Computer Science, University of Toronto.

Altenberg, B. (1991) A bibliography of publications relating to English computer corpora. English Computer corpora: selected papers and research guide, ed. by S. Johansson & A-B. Stenstroem, 355-396. Berlin: Mouton de Gruyter.

Blum-Kulka, S., J. House, & G. Kasper. (1989) *Cross-cultural pragmatics: requests and apologies*. Norwood, NJ: Ablex.

Chafe, W. (1980) The pear stories; cognitive, cultural and linguistic aspects of narrative production. Norwood, NJ: Ablex.

Edwards, J.A. & M.D. Lampert. (1993) Talking data: transcription and coding in discourse research. Hillsdale, NJ: L. Erlbaum.

Edwards, J.A. (1993) Survey of electronic corpora and related resources for language researchers. In Edwards & Martin 1993, pp. 263-310.

Fillmore, C.J. (1991) "Corpus linguistics" or "Computer-aided armchair linguistics". Directions in corpus linguistics, ed. by Jan Svartvik, 35-60. (Proceedings of Nobel Symposium 82, Stockholm, Sweden, 4-8 Aug 1991. Trends in Linguistics: Studies and Monographs 65.) Berlin: Mouton de Gruyter.

Garside, R., G. Leech & G. Sampson (eds.). (1987) The computational analysis of English. A corpus-based approach. London: Longman

Green, G.M. (1974) Semantics and syntactic regularity. Bloomington, IN: Indiana University Press.

Green, G.M. (1992) The universality of Gricean accounts of politeness; you gotta have wa. Ms.

Green, G.M. (1993) Rationality and Gricean inference. Cognitive Science Technical Report UIUC-BI-CS-93-09 (Language Series). Beckman Institute, University of Illinois at Urbana-Champaign.

Greenbaum, S., G. Leech, & J. Svartvik (eds.). (1980) Studies in English linguistics for Randolph Quirk. London: Longman.

Grosz, B.J. (1977) The representation and use of focus in dialog understanding. Ph.D. thesis, University of California, Berkeley.

Hill, B., S. Ide, S. Ikuta, A. Kawasaki, & T. Ogino. (1986) Universals of Linguistic Politeness. Quantitative Evidence from Japanese and American English. *Journal of Pragmatics* 10: 347--371.

Horn, L.R. (1984) Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. Georgetown University Round Table on Language and Linguistics 1984: Meaning, form, and use in context: Linguistic applications, ed. by D. Schiffrin, 11-42.

Horn, L.R. (1989) A natural history of negation. Chicago: University of Chicago Press.

Horrigan, M.K. (1977) Modelling simple dialogs. Master's thesis and Technical Report 108, Department of Computer Science, University of Toronto.

Ide, S., B. Hill, Y.M. Carnes, T. Ogino, & A. Kawasaki. (1992) The concept of politeness: An empirical study of American English and Japanese. *Politeness in Language*, ed. by R.J. Watts, S. Ide & K. Ehlich, 281--297. New York: Mouton de Gruyter.

Johansson, S. & A. Stenstroem. (1991) English Computer corpora; selected papers and research guide. Berlin: Mouton de Gruyter.

Meijs, W. (ed.). (1987) Corpus linguistics and beyond. Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora. Amsterdam: Rodopi.

Merritt, M. (1976) On questions following questions (in service encounters). Language in society 5:315-357.

Morgan, J.L. (1978) Two types of convention in indirect speech acts. *Syntax and semantics, Vol. 9: Pragmatics*, ed. by P. Cole, 261-280. New York: Academic Press.

Nunberg, G.D. (1978) *The pragmatics of reference*. Bloomington, IN: Indiana University Linguistics Club.

Postal, P.M. (1974) On raising. Cambridge, MA: MIT Press.

Rose, K.R. (1992) Method and scope in cross cultural speech act research: a contrastive study of requests in Japanese and English. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Santorini, B. (1990) Selections from the annotation manual for the Penn Treebank project for the ACL tutorial on tagging linguistic analysis. Ms.

Searle, J. (1969) Speech acts. Cambridge University Press.

Svartvik, J. (1990) The London-Lund Corpus of Spoken English: Description and research. Lund Studies in English 82. Lund: Lund University Press.

Terkel, S. (1972) Working. New York: Avon Books.

Terkel, S. (1967) Division Street. New York: Avon Books.

Tottie, G., & I. Baecklund (eds.). (1986) English in speech and writing. Studia Anglistica Upsaliensia 60. Stockholm: Almqvist & Wiksell.